Robust Linear Transceiver Design for Multi-Hop Non-Regenerative MIMO Relaying Systems

Chengwen Xing[†], Zesong Fei[†], Shaodan Ma[‡], Jingming Kuang[†], and Yik-Chung Wu[‡]

[†]School of Information and Electronics, Beijing Institute of Technology, Beijing, China
Email: {xingchengwen}@gmail.com {zesongfei, jmkuang}@bit.edu.cn
[‡]Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong
Email: {sdma,ycwu}@eee.hku.hk

Abstract—In this paper, optimal linear transceiver designs for multi-hop amplify-and-forward (AF) Multiple-input Multiple-out (MIMO) relaying systems with Gaussian distributed channel estimation errors are investigated. Some commonly used transceiver design criteria are unified into a single matrix-variate optimization problem. With novel applications of majorization theory and properties of matrix-variate function, the optimal structure of robust transceiver is first derived. Based on the optimal structure, the original transceiver design problems are reduced to much simpler problems with only scalar variables whose solutions are readily obtained by iterative water-filling algorithms. The performance advantages of the proposed robust designs are demonstrated by the simulation results.

I. INTRODUCTION

In order to satisfy the emerging requirements for high speed ubiquitous wireless communications, MIMO cooperative communication has become one of the key parts in the future wireless standards such as LTE, IMT-Advanced, Winner project, etc. Transceiver design for amplify-and-forward (AF) MIMO relaying systems has been reported in [1]–[6]. There are various design criteria with different goals. The most common criteria are capacity maximization [1], [2], [5] and data mean-square-error (MSE) minimization [3]–[5]. In most of the previous works on transceiver design, most of the designs are restricted for dual-hop relaying systems and furthermore channel state information (CSI) is assumed to be perfectly known. Unfortunately, channel estimation errors are inevitable in practical systems. To mitigate the effect on the performance of AF relaying systems, such channel estimation errors should be taken into account in the transceiver design process.

In this paper, we consider robust transceiver design for a **multi-hop** AF relaying system with channel estimation errors. Taking the Gaussian distributed channel errors into account, the precoder at source, multiple forwarding matrices at all the relays and equalizer at destination are jointly designed. The structure of the optimal solution for the unified problem is derived based on Majorization theory and properties of vector-monotone functions. The derived optimal structure covers most of the existing transceiver design results in point-to-point and dual-hop AF MIMO relaying systems as special cases. With the optimal structure, iterative water-filling solutions are proposed to obtain the remaining unknown parameters in the transceiver. The performance advantages of the proposed robust designs are demonstrated by simulation results.

The following notations are used throughout this paper. Boldface lowercase letters denote vectors, while boldface uppercase letters denote matrices. The notation $\mathbf{Z}^{\mathrm{H}}$ denotes the Hermitian of the matrix $\mathbf{Z}$, and $\mathrm{Tr}(\mathbf{Z})$ is the trace of the matrix $\mathbf{Z}$. The symbol $\mathbf{I}_M$ denotes the $M \times M$ identity matrix, while $\mathbf{0}_{M,N}$ denotes the $M \times N$ all zero matrix. The notation $\mathbf{Z}^{1/2}$ is the Hermitian square root of the positive semidefinite matrix $\mathbf{Z}$, such that $\mathbf{Z}^{1/2}\mathbf{Z}^{1/2} = \mathbf{Z}$ and $\mathbf{Z}^{1/2}$ is also a Hermitian matrix. The symbol $\lambda_i(\mathbf{Z})$ represents the ith largest eigenvalue of $\mathbf{Z}$. The symbol $\otimes$ denotes the Kronecker product. For two Hermitian matrices, $\mathbf{C} \succeq \mathbf{D}$ means that $\mathbf{C} - \mathbf{D}$ is a positive semi-definite matrix. For two vectors, $\mathbf{x} \geq \mathbf{y}$ represents that each element of $\mathbf{x}$ is larger than the corresponding counterpart of $\mathbf{y}$. The symbol $\mathbf{\Lambda} \searrow$ represents a rectangular diagonal matrix with decreasing diagonal elements.

II. SYSTEM MODEL

In this paper, a multi-hop AF MIMO relaying system is considered. There is one source with N_1 antennas wants to communicate with the destination with M_K antennas through $K - 1$ relays. For the kth relay, it has M_k receive antennas and N_{k+1} transmit antennas. It is obvious that the dual-hop AF MIMO relaying systems is one of its special cases when $K = 2$.

At the source, a $N \times 1$ data vector s with covariance matrix $\mathbf{R_s} = \mathbb{E}\{\mathbf{ss}^{\mathrm{H}}\} = \mathbf{I}_N$ is transmitted through a precoder matrix $\mathbf{P}_1$. The received signal $\mathbf{x}_1$ at the first relay is

$$\mathbf{x}_1 = \mathbf{H}_1\mathbf{P}_1\mathbf{s} + \mathbf{n}_1 \tag{1}$$

where $\mathbf{H}_1$ is the MIMO channel matrix between the source and the first relay, and $\mathbf{n}_1$ is the additive Gaussian noise vector at the first relay with zero mean and covariance matrix $\mathbf{R}_{n_1} = \sigma_1^2\mathbf{I}_{M_1}$.

At the first relay, the received signal $\mathbf{x}_1$ is first multiplied by a forwarding matrix $\mathbf{P}_2$ and then the resultant signal is transmitted to the second relay. The received signal $\mathbf{x}_2$ at the second relay is given by

$$\begin{aligned}\mathbf{x}_2 &= \mathbf{H}_2\mathbf{P}_2\mathbf{x}_1 + \mathbf{n}_2 \\ &= \mathbf{H}_2\mathbf{P}_2\mathbf{H}_1\mathbf{P}_1\mathbf{s} + \mathbf{H}_2\mathbf{P}_2\mathbf{n}_1 + \mathbf{n}_2,\end{aligned} \tag{2}$$

the transceiver. The performance advantages of the proposed robust designs are demonstrated by simulation results.

where $\mathbf{H}_2$ is the MIMO channel matrix between the first relay and the second relay, and $\mathbf{n}_2$ is the additive Gaussian noise vector at the second relay with zero mean and covariance matrix $\mathbf{R}_{n_2} = \sigma_2^2 \mathbf{I}_{M_2}$. Similarly, the received signal at k^{th} relay can be written as

$$\mathbf{x}_k = \mathbf{H}_k \mathbf{P}_k \mathbf{x}_{k-1} + \mathbf{n}_k \qquad (3)$$

where $\mathbf{H}_k$ is the channel for the kth hop, and $\mathbf{n}_k$ is the additive Gaussian noise with zero mean and covariance matrix $\mathbf{R}_{n_k} = \sigma_k^2 \mathbf{I}_{M_k}$.

Finally, for a K-hop AF MIMO relaying system, the received signal at the destination is

$$\mathbf{y} = \left[\prod_{k=1}^{K} \mathbf{H}_k \mathbf{P}_k \right] \mathbf{s} + \sum_{k=1}^{K-1} \left\{ \left[\prod_{l=k+1}^{K} \mathbf{H}_l \mathbf{P}_l \right] \mathbf{n}_k \right\} + \mathbf{n}_K, \quad (4)$$

where $\prod_{k=1}^{K} \mathbf{Z}_k$ denotes $\mathbf{Z}_K \times \cdots \times \mathbf{Z}_1$. In order to guarantee the transmitted data $\mathbf{s}$ can be recovered at the destination, it is assumed that N_k and M_k are greater than or equal to N [3].

In practical systems, because of limited length of training sequences, channel estimation errors are inevitable. With channel estimation errors, we can write

$$\mathbf{H}_k = \bar{\mathbf{H}}_k + \Delta \mathbf{H}_k, \qquad (5)$$

where $\bar{\mathbf{H}}_k$ is the estimated channel in the kth hop and $\Delta \mathbf{H}_k$ is the corresponding channel estimation error whose elements are zero mean Gaussian random variables. Moreover, the $M_k \times N_k$ matrix $\Delta \mathbf{H}_k$ can be decomposed using the widely used Kronecker model $\Delta \mathbf{H}_k = \Sigma_k^{1/2} \mathbf{H}_{W,k} \Psi_k^{1/2}$ [7], [8]. The elements of the $M_k \times N_k$ matrix $\mathbf{H}_{W,k}$ are independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and unit variance. The specific formulas of the row correlation matrix Σ_k and the column correlation matrix Ψ_k are determined by the training sequences and channel estimators being used [7], [8].

At the destination, a linear equalizer $\mathbf{G}$ is employed to detect the desired data vector $\mathbf{s}$. The resulting data MSE matrix equals to $\Phi(\mathbf{G}) = \mathbb{E}\{(\mathbf{G}\mathbf{y} - \mathbf{s})(\mathbf{G}\mathbf{y} - \mathbf{s})^{\text{H}}\}$, where the expectation is taken with respect to random data, channel estimation errors, and noise. Following a similar derivation in dual-hop systems [8], the MSE matrix is derived to be

$$\Phi(\mathbf{G})$$
$$= \mathbb{E}\{(\mathbf{G}\mathbf{y} - \mathbf{s})(\mathbf{G}\mathbf{y} - \mathbf{s})^{\text{H}}\}$$
$$= \mathbf{G}[\bar{\mathbf{H}}_K \mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^{\text{H}} \bar{\mathbf{H}}_K^{\text{H}} + \text{Tr}(\mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^{\text{H}} \Psi_K) \Sigma_K$$
$$+ \mathbf{R}_{n_K}] \mathbf{G}^{\text{H}} + \mathbf{I}_N - \left[\prod_{k=1}^{K} \bar{\mathbf{H}}_k \mathbf{P}_k \right]^{\text{H}} \mathbf{G}^{\text{H}} - \mathbf{G} \left[\prod_{k=1}^{K} \bar{\mathbf{H}}_k \mathbf{P}_k \right], \qquad (6)$$

where the received signal covariance matrix $\mathbf{R}_{\mathbf{x}_k}$ at the kth relay satisfies the following recursive formula

$$\mathbf{R}_{\mathbf{x}_k} = \bar{\mathbf{H}}_k \mathbf{P}_k \mathbf{R}_{\mathbf{x}_{k-1}} \mathbf{P}_k^{\text{H}} \bar{\mathbf{H}}_k^{\text{H}} + \text{Tr}(\mathbf{P}_k \mathbf{R}_{\mathbf{x}_{k-1}} \mathbf{P}_k^{\text{H}} \Psi_k) \Sigma_k + \mathbf{R}_{n_k}, \qquad (7)$$

and $\mathbf{R}_{\mathbf{x_0}} = \mathbf{R}_{\mathbf{s}} = \mathbf{I}_N$ represents the signal covariance matrix at the source.

III. TRANSCEIVER DESIGN PROBLEMS

A. Objective Functions

There are various performance metrics for transceiver designs. In the following, we focus on two widely used metrics.

(1) In general, for balancing the performance across different data streams, (e.g., minimizing the worst data stream MSE), the objective function is written as [9]

$$\text{Obj 1: } \psi_1[\text{d}(\Phi(\mathbf{G}))] \qquad (8)$$

where $\psi_1(\bullet)$ is an increasing Schur-convex function[1] and $\text{d}(\Phi(\mathbf{G})) = [[\Phi(\mathbf{G})]_{1,1} \ [\Phi(\mathbf{G})]_{2,2} \ \cdots]^{\text{T}}$, with the symbol $[\mathbf{Z}]_{i,j}$ represents the $(i,j)^{\text{th}}$ entry of $\mathbf{Z}$.

(2) On the other hand, if a preference is given over a certain data streams, (e.g., loading more resources to the data streams with better channel state information), the objective function can be written as

$$\text{Obj 2: } \psi_2[\text{d}(\Phi(\mathbf{G}))] \qquad (9)$$

where $\psi_2(\bullet)$ is an increasing Schur-concave function.

B. Problem Formulation

Although the above two criteria aim at different designs, the transceiver design optimization problem can be unified into a single form:

$$\min_{\mathbf{P}_k, \mathbf{G}} \quad \boldsymbol{f}(\Phi(\mathbf{G}))$$
$$\text{s.t.} \quad \text{Tr}(\mathbf{P}_k \mathbf{R}_{\mathbf{x}_{k-1}} \mathbf{P}_k^{\text{H}}) \leq P_k, \quad k = 1, \cdots, K \qquad (10)$$

where the objective function $\boldsymbol{f}(\bullet)$ is a real-valued matrix-variate function with $\Phi(\mathbf{G})$ as its argument. Notice that for all the two objectives described above, $\boldsymbol{f}(\bullet)$ is a matrix-monotone increasing function.

For (10), there is no constraint on the equalizer $\mathbf{G}$. We can differentiate the trace of (6) with respect to $\mathbf{G}$ and obtain the LMMSE equalizer

$$\mathbf{G}_{\text{LMMSE}} = \left[\prod_{k=1}^{K} \bar{\mathbf{H}}_k \mathbf{P}_k \right]^{\text{H}} [\bar{\mathbf{H}}_K \mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^{\text{H}} \bar{\mathbf{H}}_K^{\text{H}}$$
$$+ \text{Tr}(\mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^{\text{H}} \Psi_K) \Sigma_K + \mathbf{R}_{n_K}]^{-1}, \quad (11)$$

with the property [10]

$$\Phi(\mathbf{G}_{\text{LMMSE}}) \preceq \Phi(\mathbf{G}). \qquad (12)$$

Because $\boldsymbol{f}(\bullet)$ is a matrix-monotone increasing function, (12) implies that $\mathbf{G}_{\text{LMMSE}}$ minimizes the objective function in (10). Substituting the optimal equalizer of (11) into $\Phi(\mathbf{G})$ in (6), $\Phi(\mathbf{G})$ equals to

$$\Phi_{\text{MSE}} = \mathbf{I}_N - \left[\prod_{k=1}^{K} \bar{\mathbf{H}}_k \mathbf{P}_k \right]^{\text{H}} [\bar{\mathbf{H}}_K \mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^{\text{H}} \bar{\mathbf{H}}_K^{\text{H}}$$
$$+ \text{Tr}(\mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^{\text{H}} \Psi_K) \Sigma_K + \mathbf{R}_{n_K}]^{-1} \left[\prod_{k=1}^{K} \bar{\mathbf{H}}_k \mathbf{P}_k \right]. \qquad (13)$$

[1]The detailed introduction of Schur-concave/convex functions, and majorization theory is given in [11].

For multi-hop AF MIMO relaying systems, the received signal at k^{th} relay depends on the forwarding matrices at all preceding relays, making the power allocations at different relays couples with each other (as seen in the constrains of (10)), and thus the problem (10) difficult to solve. In order to simplify the problem, we define the following new variable in terms of $\mathbf{P}_k$:

$$
\begin{aligned}
\mathbf{F}_k \triangleq \mathbf{P}_k \mathbf{K}_{\mathbf{F}_{k-1}}^{1/2} \\
\times \underbrace{(\mathbf{K}_{\mathbf{F}_{k-1}}^{-1/2} \bar{\mathbf{H}}_{k-1} \mathbf{F}_{k-1} \mathbf{F}_{k-1}^{\text{H}} \bar{\mathbf{H}}_{k-1}^{\text{H}} \mathbf{K}_{\mathbf{F}_{k-1}}^{-1/2} + \mathbf{I}_{M_{k-1}})^{1/2}}_{\triangleq \mathbf{\Pi}_{k-1}} \mathbf{Q}_{k-1}^{\text{H}},
\end{aligned}
$$
(14)

where $\mathbf{K}_{\mathbf{F}_k} \triangleq \text{Tr}(\mathbf{F}_k \mathbf{F}_k^{\text{H}} \boldsymbol{\Psi}_k) \boldsymbol{\Sigma}_k + \sigma_{n_k}^2 \mathbf{I}_{M_k}$ and $\mathbf{Q}_k$ is an unknown unitary matrix. The introduction of $\mathbf{Q}_k$ is due to that fact that for a positive semi-definite matrix $\mathbf{M}$, its square roots has the form $\mathbf{M}^{1/2}\mathbf{Q}$ where $\mathbf{Q}$ is an unitary matrix. Notice that $\mathbf{F}_1 = \mathbf{P}_1$. With the new variable, the MSE matrix $\boldsymbol{\Phi}_{\text{MSE}}$ is reformulated as

$$
\begin{aligned}
\boldsymbol{\Phi}_{\text{MSE}} = \mathbf{I}_N - \left[\prod_{k=1}^K \mathbf{Q}_k \mathbf{\Pi}_k^{-1/2} \mathbf{K}_{\mathbf{F}_k}^{-1/2} \bar{\mathbf{H}}_k \mathbf{F}_k \right]^{\text{H}} \\
\times \left[\prod_{k=1}^K \underbrace{\mathbf{Q}_k \mathbf{\Pi}_k^{-1/2} \mathbf{K}_{\mathbf{F}_k}^{-1/2} \bar{\mathbf{H}}_k \mathbf{F}_k}_{\triangleq \mathbf{A}_k} \right] \\
= \mathbf{I}_N - \mathbf{A}_1^{\text{H}} \cdots \mathbf{A}_K^{\text{H}} \mathbf{A}_K \cdots \mathbf{A}_1.
\end{aligned}
$$
(15)

Meanwhile, with the new variables $\mathbf{F}_k$, the corresponding power constraint in the kth hop can now be rewritten as

$$
\text{Tr}(\mathbf{F}_k \mathbf{F}_k^{\text{H}}) \le P_k.
$$
(16)

It is obvious that with the new variables $\mathbf{F}_k$, the constraints become independent of each other. Putting (15) and (16) into (10), the transceiver design problem can be reformulated as

$$
\begin{aligned}
&\textbf{P 1:} \quad \min_{\mathbf{F}_k, \mathbf{Q}_k} \quad \boldsymbol{f}(\mathbf{I}_N - \boldsymbol{\Theta}) \\
&\text{s.t.} \quad \text{Tr}(\mathbf{F}_k \mathbf{F}_k^{\text{H}}) \le P_k, \quad k = 1, \cdots, K \\
&\quad\quad \boldsymbol{\Theta} = \mathbf{A}_1^{\text{H}} \cdots \mathbf{A}_K^{\text{H}} \mathbf{A}_K \cdots \mathbf{A}_1 \\
&\quad\quad \mathbf{Q}_k^{\text{H}} \mathbf{Q}_k = \mathbf{I}_{M_k}
\end{aligned}
$$
(17)

From the definition of $\mathbf{A}_k$ in (15) and noticing that $\mathbf{K}_{\mathbf{F}_k} = \text{Tr}(\mathbf{F}_k \mathbf{F}_k^{\text{H}} \boldsymbol{\Psi}_k) \boldsymbol{\Sigma}_k + \sigma_{n_k}^2 \mathbf{I}_{M_k}$, it can be seen that $\mathbf{F}_k$ appears at multiple positions in the objective function. Therefore, the optimization problem is much more complicated than the counterpart with prefect CSI. Indeed, as demonstrated by existing works, robust transceiver design for point-to-point or dual-hop relaying MIMO systems is much more complicated and challenging than its counterpart with perfect CSI [7], [8].

IV. OPTIMAL STRUCTURE OF ROBUST TRANSCEIVER

Based on the formulations of the objectives given in (8) and (9), in Appendix A, it is proved that **P 1** has the following property.

Property 1: At the optimal value of **P 1**, $\boldsymbol{\Theta}$ must have the structure of

$$
\boldsymbol{\Theta} = \mathbf{U}_{\boldsymbol{\Omega}} \text{diag}[\boldsymbol{\lambda}(\boldsymbol{\Theta})] \mathbf{U}_{\boldsymbol{\Omega}}^{\text{H}}
$$
(18)

where the vector $\boldsymbol{\lambda}(\boldsymbol{\Theta}) = [\lambda_1(\boldsymbol{\Theta}), \cdots, \lambda_N(\boldsymbol{\Theta})]^{\text{T}}$ with $\lambda_n(\boldsymbol{\Theta})$ being the nth largest eigenvalue of $\boldsymbol{\Theta}$, and

$$
\mathbf{U}_{\boldsymbol{\Omega}} = \begin{cases} \mathbf{Q}_{\mathbf{F}} & \text{for Obj 1} \\ \mathbf{I}_N & \text{for Obj 2} \end{cases} .
$$
(19)

In (19), the unitary matrix $\mathbf{U}_{\mathbf{W}}$ is defined from the eigen-decomposition $\mathbf{W} = \mathbf{U}_{\mathbf{W}} \boldsymbol{\Lambda}_{\mathbf{W}} \mathbf{U}_{\mathbf{W}}^{\text{H}}$ with $\boldsymbol{\Lambda}_{\mathbf{W}} \searrow$, the matrix $\mathbf{U}_{\text{Arb}}$ is an arbitrary unitary matrix, and $\mathbf{Q}_{\mathbf{F}}$ is the unitary matrix which makes $\mathbf{Q}_{\mathbf{F}} \text{diag}[\boldsymbol{\lambda}(\boldsymbol{\Theta})] \mathbf{Q}_{\mathbf{F}}^{\text{H}}$ having identical diagonal elements. Furthermore, with this optimal structure, the objective function of **P 1** equals to

$$
\boldsymbol{f}(\mathbf{I}_N - \underbrace{\mathbf{U}_{\boldsymbol{\Omega}} \text{diag}[\boldsymbol{\lambda}(\boldsymbol{\Theta})] \mathbf{U}_{\boldsymbol{\Omega}}^{\text{H}}}_{\boldsymbol{\Theta}}) = \boldsymbol{g}[\boldsymbol{\lambda}(\boldsymbol{\Theta})]
$$
(20)

where $\boldsymbol{g}(\bullet)$ is a monotobically decreasing and Schur-concave function with respective to $\boldsymbol{\lambda}(\boldsymbol{\Theta})$. [2]
Proof: See Appendix A. ∎

Based on **Property 1**, the objective function of (17) can be directly replaced by $\boldsymbol{g}[\boldsymbol{\lambda}(\boldsymbol{\Theta})]$ and thus the optimization problem is simplified as

$$
\begin{aligned}
&\textbf{P 2:} \quad \min_{\mathbf{F}_k, \mathbf{Q}_k} \quad \boldsymbol{g}[\boldsymbol{\lambda}(\boldsymbol{\Theta})] \\
&\quad\quad \text{s.t.} \quad \boldsymbol{\Theta} = \mathbf{A}_1^{\text{H}} \cdots \mathbf{A}_K^{\text{H}} \mathbf{A}_K \cdots \mathbf{A}_1 \\
&\quad\quad\quad \text{Tr}(\mathbf{F}_k \mathbf{F}_k^{\text{H}}) \le P_k, \quad \mathbf{Q}_k^{\text{H}} \mathbf{Q}_k = \mathbf{I}_{M_k} \\
&\quad\quad\quad \boldsymbol{\Theta} = \mathbf{U}_{\boldsymbol{\Omega}} \text{diag}[\boldsymbol{\lambda}(\boldsymbol{\Theta})] \mathbf{U}_{\boldsymbol{\Omega}}^{\text{H}}
\end{aligned}
$$
(21)

where $\mathbf{A}_k$'s are defined in (15). In order to further simplify the optimization problem, we make use of the following two additional properties.
Property 2: As $\boldsymbol{g}(\bullet)$ is a decreasing and Schur-concave function and $\boldsymbol{\lambda}(\boldsymbol{\Theta}) \prec_w \boldsymbol{\gamma}(\boldsymbol{\Theta})$, the objective function in **P 2** satisfies

$$
\boldsymbol{g}(\boldsymbol{\lambda}(\boldsymbol{\Theta})) \ge \boldsymbol{g}(\underbrace{[\gamma_1(\boldsymbol{\Theta}) \cdots \gamma_N(\boldsymbol{\Theta})]^{\text{T}}}_{\triangleq \boldsymbol{\gamma}(\boldsymbol{\Theta})})
$$
(22)

with $\gamma_i(\boldsymbol{\Theta}) \triangleq \lambda_i(\mathbf{A}_K^{\text{H}} \mathbf{A}_K) \lambda_i(\mathbf{A}_{K-1}^{\text{H}} \mathbf{A}_{K-1}) \cdots \lambda_i(\mathbf{A}_1^{\text{H}} \mathbf{A}_1)$,
(23)

where the equality in (22) holds when the neighboring $\mathbf{A}_k$'s satisfy

$$
\mathbf{V}_{\mathbf{A}_k} = \mathbf{U}_{\mathbf{A}_{k-1}}, \quad k = 2, \cdots, K
$$
(24)

with unitary matrices $\mathbf{U}_{\mathbf{A}_k}$ and $\mathbf{V}_{\mathbf{A}_k}$ being defined based on the following singular value decomposition $\mathbf{A}_k = \mathbf{U}_{\mathbf{A}_k} \boldsymbol{\Lambda}_{\mathbf{A}_k} \mathbf{V}_{\mathbf{A}_k}^{\text{H}}$ with $\boldsymbol{\Lambda}_{\mathbf{A}_k} \searrow$.
Property 3: As $\boldsymbol{g}(\bullet)$ is a monotonically decreasing function with respective to its vector argument, the optimal solutions of the optimization problem always occur on the boundary:

$$
\text{Tr}(\mathbf{F}_k \mathbf{F}_k^{\text{H}}) = P_k.
$$
(25)

[2]The specific expressions of $\boldsymbol{g}(\bullet)$ are given in Appendix A, but they are not important for the derivation of the optimal structures.

Furthermore, defining

$$\eta_{f_k} \triangleq \mathrm{Tr}(\mathbf{F}_k\mathbf{F}_k^{\mathrm{H}}\boldsymbol{\Psi}_k)\alpha_k + \sigma_{n_k}^2 \qquad (26)$$

with $\alpha_k = \mathrm{Tr}(\boldsymbol{\Sigma}_k)/M_k$ which is a constant, (25) is equivalent to

$$\mathrm{Tr}[\mathbf{F}_k\mathbf{F}_k^{\mathrm{H}}(\alpha_k P_k\boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_k})]/\eta_{f_k} = P_k. \qquad (27)$$

Based on **Properties 2** and **3**, the optimal solution of the optimization problem (21) is exactly the optimal solution of the following new optimization problem with different constraints

P 3: $\min\limits_{\mathbf{F}_k,\mathbf{Q}_k} \quad \boldsymbol{g}[\boldsymbol{\gamma}(\boldsymbol{\Theta})]$

$$\text{s.t.} \quad \mathrm{Tr}[\mathbf{F}_k\mathbf{F}_k^{\mathrm{H}}(\alpha_k P_k\boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_k})]/\eta_{f_k} = P_k$$
$$\boldsymbol{\Theta} = \mathbf{A}_1^{\mathrm{H}}\cdots\mathbf{A}_K^{\mathrm{H}}\mathbf{A}_K\cdots\mathbf{A}_1$$
$$\mathbf{Q}_k^{\mathrm{H}}\mathbf{Q}_k = \mathbf{I}_{M_k}, \quad \boldsymbol{\Theta} = \mathbf{U}_{\boldsymbol{\Omega}}\mathrm{diag}[\boldsymbol{\gamma}(\boldsymbol{\Theta})]\mathbf{U}_{\boldsymbol{\Omega}}^{\mathrm{H}}$$
$$\mathbf{V}_{\mathbf{A}_k} = \mathbf{U}_{\mathbf{A}_{k-1}}, \quad k = 2,\cdots,K. \qquad (28)$$

Noticing that $\boldsymbol{g}(\bullet)$ is a monotonically decreasing function, solving **P 3** gives the following structure for the optimal solution.

Conclusion 1: Defining unitary matrices $\mathbf{U}_{\mathcal{H}_k}$ and $\mathbf{V}_{\mathcal{H}_k}$ based on the following singular value decomposition

$$(\mathbf{K}_{\mathbf{F}_k}/\eta_{f_k})^{-1/2}\bar{\mathbf{H}}_k(\alpha_k P_k\boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_k})^{-1/2} = \mathbf{U}_{\mathcal{H}_k}\boldsymbol{\Lambda}_{\mathcal{H}_k}\mathbf{V}_{\mathcal{H}_k}^{\mathrm{H}}$$

with $\boldsymbol{\Lambda}_{\mathcal{H}_k}\searrow$ and $\mathbf{U}_{\mathcal{H}_0} = \mathbf{U}_{\boldsymbol{\Omega}}$, $\qquad (29)$

when $\boldsymbol{\Psi}_k \propto \mathbf{I}$ or $\boldsymbol{\Sigma}_k \propto \mathbf{I}$, the optimal solutions of the optimization problem (28) have the following structure

$$\mathbf{F}_{k,\mathrm{opt}} = \sqrt{\boldsymbol{\xi}_k(\boldsymbol{\Lambda}_{\mathcal{F}_k})}(\alpha_k P_k\boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_k})^{-1/2}$$
$$\times \mathbf{V}_{\mathcal{H}_k,N}\boldsymbol{\Lambda}_{\mathcal{F}_k}\mathbf{U}_{\mathcal{H}_{k-1},N}^{\mathrm{H}}$$
$$\mathbf{Q}_{k,\mathrm{opt}} = \mathbf{I}_{M_k}, \qquad (30)$$

where $\mathbf{V}_{\mathcal{H}_k,N}$ and $\mathbf{U}_{\mathcal{H}_k,N}$ are the matrices consisting of the first N columns of $\mathbf{V}_{\mathcal{H}_k}$ and $\mathbf{U}_{\mathcal{H}_k}$, respectively, and $\boldsymbol{\Lambda}_{\mathcal{F}_k}$ is a $N \times N$ unknown diagonal matrix. The scalar $\boldsymbol{\xi}_k(\boldsymbol{\Lambda}_{\mathcal{F}_k})$ is a function of $\boldsymbol{\Lambda}_{\mathcal{F}_k}$ and equals to

$$\boldsymbol{\xi}_k(\boldsymbol{\Lambda}_{\mathcal{F}_k}) = \eta_{f_k}$$
$$= \sigma_{n_k}^2/\{1 - \alpha_k\mathrm{Tr}[\mathbf{V}_{\mathcal{H}_k,N}^{\mathrm{H}}(\alpha_k P_k\boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_k})^{-1/2}$$
$$\times \boldsymbol{\Psi}_k(\alpha_k P_k\boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_k})^{-1/2}\mathbf{V}_{\mathcal{H}_k,N}\boldsymbol{\Lambda}_{\mathcal{F}_k}^2]\}. \qquad (31)$$

In the optimal structure given by (30), the scalar variable $\boldsymbol{\xi}_k(\boldsymbol{\Lambda}_{\mathcal{F}_k})$ is only a function of the matrix $\boldsymbol{\Lambda}_{\mathcal{F}_k}$ and therefore the only unknown variable in (30) is $\boldsymbol{\Lambda}_{\mathcal{F}_k}$. The remaining unknown diagonal elements of $\boldsymbol{\Lambda}_{\mathcal{F}_k}$ can be obtained by water-filling alike solution as discussed in the next section.

V. COMPUTATIONS OF $\boldsymbol{\Lambda}_{\mathcal{F}_k}$

The remaining unknown variables in (30) are only $\boldsymbol{\Lambda}_{\mathcal{F}_k}$. Substituting the optimal structures given by **Conclusion 1** into **P 3** and defining $[\boldsymbol{\Lambda}_{\mathcal{H}_k}]_{i,i} = h_{k,i}$ and $[\boldsymbol{\Lambda}_{\mathcal{F}_k}]_{i,i} = f_{k,i}$ for $i = 1,\cdots,N$, the optimization problem for computing $\boldsymbol{\Lambda}_{\mathcal{F}_k}$ becomes

$$\min\limits_{f_{k,i}} \quad \boldsymbol{g}[\boldsymbol{\gamma}(\boldsymbol{\Theta})]$$

$$\text{s.t.} \quad \sum_{i=1}^{N} f_{k,i}^2 = P_k$$
$$\boldsymbol{\gamma}(\boldsymbol{\Theta}) = [\gamma_1(\boldsymbol{\Theta})\ \cdots\gamma_N(\boldsymbol{\Theta})]^{\mathrm{T}}$$
$$\gamma_i(\boldsymbol{\Theta}) = \frac{\prod_{k=1}^{K} f_{k,i}^2 h_{k,i}^2}{\prod_{k=1}^{K}(f_{k,i}^2 h_{k,i}^2 + 1)}. \qquad (32)$$

The specific methods for finding $f_{k,i}$ depend on the expressions of $\boldsymbol{g}(\bullet)$, In the following, we discuss the solution of (32) in more detail. The design criterion of MAX-MSE minimization is taken as example to show how to compute $\boldsymbol{\Lambda}_{\mathcal{F}_k}$.

MAX-MSE minimization is a special case of **Obj 1** in (8) and in this case, $\boldsymbol{\psi}_1(\mathrm{d}(\boldsymbol{\Phi}_{\mathrm{MSE}})) = \max[\boldsymbol{\Phi}_{\mathrm{MSE}}]_{i,i}$. Furthermore, in Appendix A it is proved that $\boldsymbol{g}(\boldsymbol{\lambda}(\boldsymbol{\Theta})) = \boldsymbol{\psi}_1[\mathbf{1}_N - (\sum_{i=1}^{N}\lambda_i(\boldsymbol{\Theta})/N) \otimes \mathbf{1}_N]$. Therefore, $\boldsymbol{g}[\boldsymbol{\gamma}(\boldsymbol{\Theta})]$ equals to

$$\boldsymbol{g}[\boldsymbol{\gamma}(\boldsymbol{\Theta})] = \max\left(\mathbf{1}_N - \left(\sum\nolimits_{i=1}^{N}\gamma_i(\boldsymbol{\Theta})/N\right) \otimes \mathbf{1}_N\right)$$
$$= 1 - \frac{1}{N}\sum_{i=1}^{N}\gamma_i(\boldsymbol{\Theta}) \qquad (33)$$

based on which the optimization problem (32) becomes

$$\min\limits_{f_{k,i}} \quad 1 - \frac{1}{N}\sum_{i=1}^{N}\frac{\prod_{k=1}^{K}(f_{k,i}^2 h_{k,i}^2)}{\prod_{k=1}^{K}(f_{k,i}^2 h_{k,i}^2 + 1)}$$
$$\text{s.t.} \quad \sum_{i=1}^{N} f_{k,i}^2 = P_k. \qquad (34)$$

The optimization problem (34) can be solved by using iterative water-filling algorithm.

VI. SIMULATION RESULTS AND DISCUSSIONS

In this section, the performance of the proposed robust designs are evaluated by simulations. For the purpose of comparison, the algorithms based on the estimated channel only (without taking the channel estimation errors into account) are also simulated. In the following, we consider a three-hop AF MIMO relaying system where all nodes are equipped with 4 antennas. Furthermore, the estimation error correlation matrices are chosen as the popular exponential model [8] i.e., $[\boldsymbol{\Psi}_k]_{i,j} = \alpha^{|i-j|}$ and $[\boldsymbol{\Sigma}_k]_{i,j} = \beta^{|i-j|}$. The estimated channels $\bar{\mathbf{H}}_k$'s, are generated based on the following complex Gaussian distributions

$$\bar{\mathbf{H}}_k \sim \mathcal{CN}_{M_k,N_k}\left(\mathbf{0}_{M_k,N_k}, \frac{(1-\sigma_e^2)}{\sigma_e^2}\boldsymbol{\Sigma}_k \otimes \boldsymbol{\Psi}_k^{\mathrm{T}}\right), \qquad (35)$$

such that channel realizations $\mathbf{H}_k = \bar{\mathbf{H}}_k + \Delta\mathbf{H}_k$ have unit variance. We define the signal-to-noise ratio (SNR) for the kth link as $P_k/\sigma_{n_k}^2$. At the source node, four independent data streams are transmitted and in each data stream, $N_{Data} =$
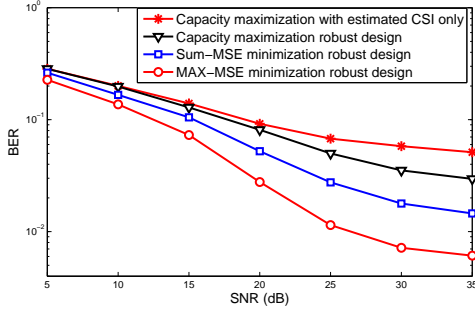
Fig. 1. BERs of the proposed robust design with different design objectives, when $\alpha = 0.4$, $\beta = 0$ and $\sigma_e^2 = 0.004$.

10000 independent QPSK symbols are transmitted. Each point in the following figure is an average of 10000 trials.

Fig. 1 shows the bit error rate (BER) of the proposed robust designs with different performance metrics: sum MSE minimization, mutual information maximization and MAX-MSE minimization. Other parameters are taken as $\alpha = 0.4$, $\beta = 0$ and $\sigma_e^2 = 0.004$. It can be seen that the former two criteria have better performance than the latter one. Meanwhile, the capacity maximization based on estimated CSI only is given to show that the proposed robust designs are better than that of the design with estimated CSI only.

VII. CONCLUSIONS

Bayesian robust transceiver design for multi-hop AF MIMO relaying systems under channel estimation errors was considered. Various transceiver design criteria were unified into a single optimization framework. Using majorization theory and properties of matrix-variate functions, the optimal structure of transceivers was derived. Then, the transceiver design problems were greatly simplified and the remaining unknowns were obtained by iterative water-filling solutions. The performance of the proposed transceiver designs has been demonstrated by simulation results.

APPENDIX A
PROOF OF PROPERTY 1

Obj 1: For the diagonal elements of the positive semi-definite matrix $\mathbf{\Phi}_{\mathrm{MSE}} = \mathbf{I}_N - \mathbf{\Theta}$, we have the following relationship [11]

$$\mathbf{1}_N - \left(\sum_{i=1}^{N} \lambda_i(\mathbf{\Theta})/N\right) \otimes \mathbf{1}_N \prec \mathrm{d}(\mathbf{I}_N - \mathbf{\Theta}) \tag{36}$$

with the equality holds if and only if $[\mathbf{\Theta}]_{i,i} = \sum_{i=1}^{N} \lambda_i(\mathbf{\Theta})/N$, where $\mathbf{1}_N$ is the $N \times 1$ all one vector.

For the first objective function in (8), as $\boldsymbol{\psi}_1(\bullet)$ is decreasing and Schur-convex, the objective function satisfies [9]

$$\boldsymbol{\psi}_1(\mathrm{d}(\mathbf{I}_N - \mathbf{\Theta})) \geq \underbrace{\boldsymbol{\psi}_1\left(\mathbf{1}_N - \left(\sum_{i=1}^{N} \lambda_i(\mathbf{\Theta})/N\right) \otimes \mathbf{1}_N\right)}_{\triangleq \boldsymbol{g}[\boldsymbol{\lambda}(\mathbf{\Theta})]},$$

$$\tag{37}$$

with equality holds if and only if $[\mathbf{\Theta}]_{i,i} = \sum_{i=1}^{N} \lambda_i(\mathbf{\Theta})/N$. Therefore, $\mathbf{\Theta}$ must have the following structure [9]

$$\mathbf{\Theta} = \mathbf{Q_F} \mathrm{diag}(\boldsymbol{\lambda}(\mathbf{\Theta})) \mathbf{Q_F^H}. \tag{38}$$

where $\mathbf{Q_F}$ is a unitary matrix such that $\mathbf{\Theta}$ has identical diagonal elements.

Based on the definition that $\boldsymbol{\psi}_1(\bullet)$ is a decreasing and Schur-convex function, based on **A.6.Lemma** and **A.8.Lemma** in [11] it can be directly proved that $\boldsymbol{g}(\boldsymbol{\lambda}(\mathbf{\Theta}))$ is a decreasing and Schur-concave function of $\boldsymbol{\lambda}(\mathbf{\Theta})$.

Obj 2: Notice that for the positive semi-definite matrix $\overline{\mathbf{\Phi}}_{\mathrm{MSE}} = \mathbf{I}_N - \mathbf{\Theta}$, $\mathrm{d}(\mathbf{I}_N - \mathbf{\Theta}) \prec \boldsymbol{\lambda}(\mathbf{I}_N - \mathbf{\Theta})$ [9]. Furthermore $\boldsymbol{\psi}_2(\bullet)$ is Schur-concave, we have

$$\boldsymbol{\psi}_2(\mathrm{d}(\mathbf{I}_N - \mathbf{\Theta})) \geq \underbrace{\boldsymbol{\psi}_2(\mathbf{1}_N - \boldsymbol{\lambda}(\mathbf{\Theta}))}_{\triangleq \boldsymbol{g}[\boldsymbol{\lambda}(\mathbf{\Theta})]}. \tag{39}$$

In order to make the equality in (39) hold, we need $[\mathbf{\Theta}]_{i,i} = \lambda_i(\mathbf{\Theta})$, which means that $\mathbf{\Theta}$ is a diagonal matrix. Therefore, we can write

$$\mathbf{\Theta} = \mathbf{I}_N \mathrm{diag}(\boldsymbol{\lambda}(\mathbf{\Theta})) \mathbf{I}_N. \tag{40}$$

Since $\boldsymbol{\psi}_2(\bullet)$ is increasing and Schur-concave, based on **A.6.Lemma** and **A.8.Lemma** in [11] it is obvious that $\boldsymbol{\psi}_2(\mathbf{1}_N - \boldsymbol{\lambda}(\mathbf{\Theta}))$ is decreasing and Schur-concave with respective to $\boldsymbol{\lambda}(\mathbf{\Theta})$.

REFERENCES

[1] O. Munoz-Medina, J. Vidal, and A. Agustin, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2593–2604, June 2007.

[2] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1398–1407, Apr. 2007.

[3] W. Guan and H. Luo, "Joint MMSE transceiver design in non-regenerative MIMO relay systems," *IEEE Commun. Lett.*, vol. 12, no. 7, pp. 517–519, July 2008.

[4] R. Mo and Y. Chew, "Precoder design for non-regenerative MIMO relay systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5041–5049, Oct. 2009.

[5] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear nonregenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4837–4851, Dec. 2009.

[6] Y. Rong and Y. Hua, "Optimality of diagonalization of multi-hop MIMO relays," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 6068–6077, Dec. 2009.

[7] C. Xing, S. Ma, Y.-C. Wu, and T.-S. Ng, "Transceiver design for dual-hop nonregenerative MIMO-OFDM relay systems under channel uncertainties," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6325–6339, Dec. 2010.

[8] C. Xing, S. Ma, and Y.-C. Wu, "Robust joint design of linear relay precoder and destination equalizer for dual-hop amplify-and-forward MIMO relay systems," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2273–2283, Apr. 2010.

[9] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.

[10] S. Kay, *Fundamental of Statistical Signal Processing: Estimation Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[11] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press, 1979.